(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 267 283 A2

EUROPEAN PATENT APPLICATION

(12)

(43) Date of publication:
18.12.2002 Bulletin 2002/51

(51) Int Cl.[7]: **G06F 17/30**

(21) Application number: **02396073.5**

(22) Date of filing: **23.05.2002**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **15.06.2001 FI 20011277**

(71) Applicant: **Nokia Corporation**
02150 Espoo (FI)

(72) Inventors:
- **Koskimies, Oskari**
  00950 Helsinki (FI)
- **Juhola, Janne**
  37550 Lempäälä (FI)

(74) Representative: **Kaukonen, Juha Veikko**
Kolster Oy Ab,
Iso Roobertinkatu 23,
P.O. Box 148
00121 Helsinki (FI)

(54) Selecting data for synchronization

(57) A method for selecting a data set to be synchronized from databases of a data system, in which system metadata illustrating the relationships between data units of the data system are stored for the selection of the data set to be synchronized. The metadata comprises at least information on the relevance between the data units. When a first data set is to be synchronized, metadata associated with at least one initial data unit of the first data set is retrieved. Next, a second data set, which according to at least one metadata element comprises a data unit of maximum relevance to the initial data unit, is selected for synchronization.

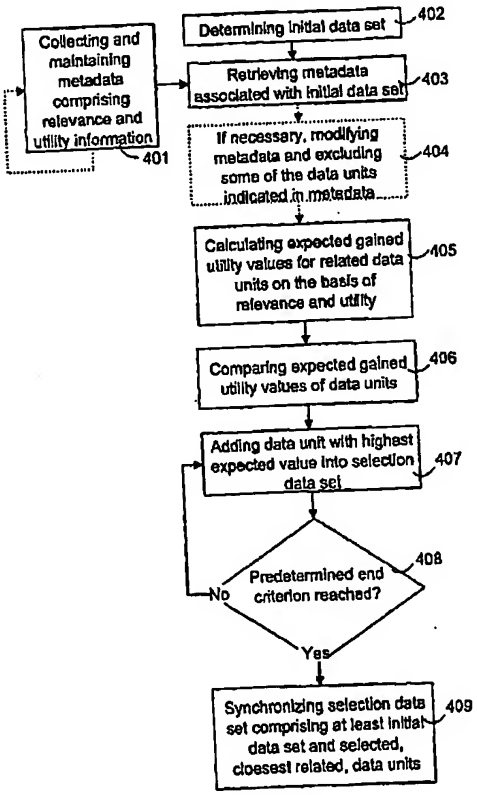

Fig. 4

EP 1 267 283 A2

Description

BACKGROUND OF THE INVENTION

[0001] The invention relates to selecting data for synchronization. Data synchronization is an operation in which a correspondence is provided between the data collections of at least two databases to the effect that, after the synchronization, the data units of the data collections substantially correspond to each other. The term 'database' should be understood in its broad sense to relate to any data collection which resides in a data source or data storage and which can be updated using one or more applications.

[0002] Along with the increasing popularity of new networking terminals, such as portable computers, PDA (Personal Digital Assistant) devices, mobile stations and pagers, the need for data synchronization has increased as well. Data of calendar and electronic mail applications in particular are typical examples of data that need to be synchronized. Synchronization has conventionally been based on different proprietary protocols, which are not compatible with each other. However, in mobile communications in particular, it is important that data can be obtained and updated irrespective of the terminal and application in use.

[0003] For improved synchronization of application data, a Synchronization Mark-up Language (SyncML) based on the Extensible Mark-up Language (XML) has been developed. A SyncML synchronization protocol employing messages of SyncML format allows the data of any application to be synchronized between any networked terminals. The SyncML synchronization protocol functions both in wireless and in fixed networks and supports a plural number of transmission protocols. SyncML provides both a synchronization protocol and a data representation protocol.

[0004] The implementation of data synchronization is described in the SyncML standard, but the standard does not specify in detail how to select the data that is to be synchronized. Typically, the amount of data on a server or desktop computer considerably exceeds the capacity of a portable device. Even larger portable terminals, such as portable computers, are not necessarily able to store all the data needed by the user, for example copies of every important document contained in a company's data system. If synchronization is carried out over the radio interface, further restrictions are caused by the available bandwidth. From the user's point of view, synchronization over the radio interface may appear to be too slow, and in a mobile communications network the transmission costs may be too high. Consequently, it is necessary to restrict the amount of data to be synchronized by selecting only a subset of the data for synchronization. This may be called 'adaptive synchronization'. However, it is not easy to select the subset. For example, when electronic mail messages are to be synchronized, subsets such as 'New Items', 'Outgoing Items' and 'Deleted Items' could be useful. However, among the New Items, there may be a message that refers to a previous one on the same subject, in which case an important message might be inaccessible to the user. The selecting of the data to be synchronized thus depends on various factors, such as the application concerned, the terminal and the needs of the user.

[0005] In the prior art, adaptive synchronization is restricted to certain application-specific techniques that simply allow specific data units to be excluded from the data to be synchronized. A typical example is to rule out electronic mail attachment files. US 6,052,735 discloses a method in which only some of the attachment files of electronic mail messages are synchronized between a computer and a wireless terminal. The synchronization may be based on the user's choice or on filtering, in which case only pre-determined attachment files will be synchronized. In that case only electronic mail messages transferred according to a specific transfer technique can be synchronized. However, US 6,052,735 does not provide a solution for efficient selection of the data to be synchronized. In addition, prior art solutions do not take into account the different needs of applications. The SyncML protocol provides a kind of an adaption possibility in which the server is aware of the restrictions of the terminal. This means that the terminal application does not need to support all fields of a data unit and the amount of data can thus be reduced. Nevertheless, all data units are still fetched to the terminal in this case, too.

BRIEF DESCRIPTION OF THE INVENTION

[0006] It is therefore an object of the invention to provide an improved method and equipment implementing the method to allow data to be selected for synchronization such that the most important data units are selected. The objectives of the invention are achieved with a method, synchronization system, synchronization device and computer software product characterized by what is stated in the independent claims. Preferred embodiments of the invention are disclosed in the dependent claims.

[0007] The invention is based on maintaining in the data system metadata on the relationships between the data units for the purpose of selecting the data to be synchronized. The metadata comprises at least information about relevance relationships between the data units. Relevance is preferably given as a numerical value to express the probability of the user needing a data unit associated with an initial data unit, either directly or through other data units, provided that the initial data unit has been selected. In the system, metadata relating to at least one initial data unit of

a first data set is retrieved when the first data set is to b synchronized. On the basis of the metadata, a second data set, comprising at least on data unit that is most relevant to th initial data unit, is selected for synchronizati n. Typically, in addition to the first data set, data units outside the first data set that are most relevant to th initial data units are s lected into th second data set. On the other hand, it is also possibl that only most relevant initial data units from the first data set are s lected into th second data s t on the basis of the metadata.

[0008] The solution of the inv ntion provides an advantage in that it allows different relationships between the data units to be taken into account for selecting the second data set to b synchronized. This allows the most relevant data units to be selected for synchronization, and thereby the restricted terminal resources and the limited bandwidth available in wireless data transmission are more efficiently utilized. Since relevant data units can be automatically selected for synchronization, the user does not need to separately define or restrict the data units to be synchronized, which provides improved usability. Since the method can be used in different applications, the relationships between the applications can be taken into account.

[0009] According to a preferred embodiment of the invention, situation-specific metadata are defined into the data system for different synchronization situations. On the basis of the synchronization situation concerned, metadata representing the relationships between the data units is selected. The synchronization situation may be defined for example in the form of profile alternatives available to the user, such as a business trip profile or a holiday trip profile. The advantage of this embodiment is that it further improves the possibilities to take the user's needs into account when data is selected for synchronization.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] In the following, the invention will be described in connection with the preferred embodiments and with reference to the accompanying drawings, in which

Figure 1 is a general view of a data system in which the data of the databases can be synchronized;
Figure 2 is a metadata graph;
Figure 3 shows a path illustrating the relationships between data units;
Figure 4 is a flow diagram illustrating a method according to a preferred embodiment of the invention;
Figure 5 is a flow diagram illustrating a method according to a second preferred embodiment of the invention; and
Figure 6 shows an initial data set and adjacent data units associated with it.

DETAILED DESCRIPTION OF THE INVENTION

[0011] Figure 1 illustrates a networked data system, in which data comprised in separate databases DB and terminals TE can be synchronized. From the point of view of synchronization, the terminal TE is a Client Device, and it is typically a portable computer, PDA device, mobile station or pager, and a synchronization server S is a server, typically serving a plurality of client devices. However, the synchronization server is not restricted to any particular equipment type; unlike in the example described, a wireless terminal can also function as a synchronization server. Figure 1 shows two examples, the first one of which comprises terminals TE, databases DB and synchronization servers S connected to a Local Area Network LAN. A terminal TE connected to the network LAN comprises a functionality, such as a network card and software controlling data transmission, for communicating with the devices in the network LAN. The local area network LAN may be a local area network of any type, and the TE may communicate with the server S also over the Internet, typically through a firewall FW.

[0012] In the second example, the terminal TE, synchronization server S and databases DB are connected to a wireless network WNW. The terminal TE connected to the network WNW comprises a mobile communications functionality for wireless communication with the network WNW. The wireless network WNW may be any already known wireless network, such as a network supporting a GSM service, a network supporting a GPRS service (General Packet Radio Service), a third generation mobile communications network, such as a UMTS network (Universal Mobile Telecommunications System), a wireless local area network WLAN, or a private network. It is to be noted that the server S may also comprise a database DB, although in Figure 1 the servers S and the databases DB are separate, for the sake of clarity.

[0013] The terminals TE (in wired networks LAN and wireless networks WNW) and the servers S comprise memory MEM; SMEM, a user interface UI; SUI, I/O means I/O; SI/O for arranging data transmission, and a Central Processing Unit CPU;SCPU comprising one or more processors. Application data that is to be synchronized may be stored in the TE memory MEM (which, from the point of view of synchronization, may be a database to be synchronized), the database DB memory, as well as the server S memory SMEM. In response to a computer program code stored in the memory MEM; SMEM and executed in the central processing units CPU and SCPU, the terminal TE and the synchronization server S execute the inventive means, some embodiments of which are shown in Figures 4 and 5. The computer

programs may be obtained through the network and/or they may be stored in memory means, such as a disc, CD-ROM disc or other external memory means from which they can be loaded into the memory MEM; SMEM. Hardware solutions can also be used.

[0014] Metadata on the relationships between the data units are maintained in the data system. Figure 2 shows an example of a metadata graph. The nodes in the graph represent the data units and the links depicted by arrows illustrate the relationships between the data units. Each link is assigned at least one value expressing how closely the target node is associated with the source node (the closeness of the relationship). The metadata graph is preferably a directional network. As shown in Figure 2, relationships between different types of data units (depicted in different shapes) are also preferably determined. A thicker link is used in Figure 2 to denote a close relationship between the data units, whereas a thinner link is used for a remote relationship. A simple metadata graph could comprise for example an electronic mail data unit linked at least with earlier electronic mail messages on the same subject, with the contact information of the sender or the recipients, and with attachment files, if any, attached to the data unit.

[0015] Synchronization requires the determining of an initial data set the data units of which are at least to be synchronized. The metadata links allow paths from the initial data set to different data units to be determined. Figure 3 illustrates paths from initial data unit A to data unit B according to a preferred embodiment of the invention. In a preferred embodiment, the relationship between the data units is denoted by relevance and utility. Relevance is a value representing the probability that the user will need a data unit associated with an initial data unit, either directly or through other data units, provided that the initial data unit has been selected. In Figure 3, relevance is denoted by *ri*. Utility expresses the utility of a data unit associated with an initial data unit in the metadata, either directly or over a link through other data units, provided that the initial data unit has been selected. Utility can be thought of as added value obtained by a related data unit, or, on the other hand, as a loss, if the data unit is not available even if it were needed. In Figure 3 utility is shown by *ui*, each link between A and B being provided with a relevance value *ri* and a utility value *ui*. The initial data unit A and the related data unit B may be connected by several paths. The different paths represent different reasons why a user who needs the initial data unit A might also need the data unit B. In Figure 3, there are two paths p1 and p2 between A and B, the paths having the following probabilities:

$$p1 = P(p1) = r1 * r2$$

$$p2 = P(p2) = r3 * r4 * r5.$$

[0016] Hence, the relevance of B to A is the product of the relevance values assigned to the data units along the path. B's utility to A is determined by the utility value of the last link, i.e. utility through path p1 is u2 and through path p2 it is u5. Gained Utility g is the utility of the data units that the user would really request. Since the user's actions cannot be known in advance, gained utility is a random variable and therefore has a distribution and expected value. The closeness of the relationship between the data units A and B, i.e. the importance of data unit B in the selection of data unit A, can be defined by calculating an Expected Gained Utility E(g) value. If the user needs the data unit B for several different reasons (a plurality of paths p1, p2), the gained utility obtained with the data unit B can be determined in the form of the maximum utility of the paths *(max(u2,u5)*. It is also possible to use the utility of an individual path or the combined utilities of different paths as the utility to be gained by data unit selection. The expected gained utility E (g) is preferably calculated by taking into account both paths p1, p2, whereby the following is obtained:

$$E(g)=u2 * P(p1) * (1-P(p2)) + u5 * P(p2) * (1- P(p1)) + max(u2,u5) * P(p1) * P(p2).$$

[0017] If the utility value *ui* of the links is set at one, the expected gained utility E(g) represents the probability of a data unit being needed for some reason. Hence, in the example of Figure 3, E(g) is

$$P(p1) * (1 - P(p2)) + P(p2) * (1 - P(p1)) + P(p1) * P(p2) =$$

$$P(p1) - P(p1) * P(p2) + P(p2) - P(p2) * P(p1) + P(p1) * P(p2) =$$

$$P(p1) + P(p2) - P(p1) * P(p2) =$$

$P(p1 \cup p2)$.

[0018] A comparison of the expected gained utility E(g) values of related data units allows th data units comprising the highest values to be select d, in addition to the initial data units, into th s lection data set that is to be synchronized. The metadata can be collected by applying a minimum spanning tree method or by means of content analysis, for example. To optimize the processing resources and th time r quired, deviations from th above calculation method can be made. For example, the number of paths to be taken into consideration can be restricted to only comprise direct links, in which case path length is one. Methods for restricting the number of the paths to be taken into account include Dijkstra's minimum path algorithm and Kruskal's algorithm.

[0019] Figure 4 shows a method according to a preferred embodiment of the invention. Metadata comprising relevance and utility information are collected 401 into the system as described above. The metadata can be maintained in the memory MEM, SMEM in data structures, in the application executing the method, or in the application input data. Metadata can also be loaded from network databases, through the Internet, for example. To the metadata is added a new initial data unit that is to be synchronized, the related data units and utility and relevance values illustrating the relationships between them. According to a preferred embodiment, general rules are used, such as: the relevance value on a link from any electronic mail item to any related word processing file is always 0.7. Consequently, the value 0.7 is always used, irrespective of the electronic mail item or the word processing file, which reduces the space needed for storing the metadata.

[0020] In a preferred embodiment, the metadata is application-specific. In that case, new metadata needed for selecting data units for a new application are added for example to application-specific directories in the synchronizing e-device (TE, S). The metadata determines the relationships between electronic mail data items synchronized by an e-mail application, for example. In other words, the metadata from which the relationships between the data units are to be fetched are selected according to the application employed. Application-specific metadata can also be used for influencing the relationships between the data units of different applications by applying different relevance and/or utility values to the links between them. For example, a link from an electronic mail item to a word processing file has a higher relevance value than a link from a calendar entry to a word processing file. Application-specific metadata can also be used in a table format, for example, in which the relevance and/or utility values between different applications are given.

[0021] Application-specific metadata can be modified according to the purpose of use, and, in addition, different metadata can be used in different situations, i.e. for different synchronization contexts. For example, when a person is leaving for a business trip, the relevance of business card data units is higher than when s/he is leaving for a holiday trip. Metadata can be arranged for use in different synchronization contexts by applying different application- or device-level user profiles, similarly as user profiles arranged at mobile stations. Profile-specific metadata may be stored for the different profiles; it is also possible to modify the metadata or to select the data units to be synchronized on the basis of different criteria in different situations. Typical synchronization contexts include a general context, business trip, holiday trip, reading of electric mail messages and meetings. For example, when a meeting has been scheduled for the user (which can be stated from the calendar), data is synchronized with the user's terminal TE such that the business cards of those participating in the meeting form the initial data set and they are provided with links of high relevance values to the electronic mail messages last sent by the participants.

[0022] The user also has the possibility to influence the metadata, for example by adding new links between the data units, or to change the utility or relevance values of the links. To maintain good usability, a predetermined number of high-level user preferences can be defined, the metadata being automatically determined and modified according to the preferences. This could be illustrated by an example in which the user considers business cards not to be important and thus selects a low priority for them. The synchronization application may therefore set low relevance values for business cards. All preferences related to synchronization can be determined user-specifically, and the appropriate preferences can be selected using the user ID (the preferences can also be stored on an Integrated Circuit (IC) card, for example).

[0023] According to yet another embodiment of the invention, metadata can be collected and updated 401 by analyzing the contents of the data units. In response to changes in the data unit contents, the relevance and/or utility values of the contents can be changed as well.

[0024] Metadata updating 401 can be arranged to take place as an automatic monitoring of user actions. This means that a new data unit with its relevance data can be automatically added to the metadata when the user requests for the data unit in question. In addition, the frequency of use of the data units can be monitored and the relevance and/or utility values changed automatically on the basis of the monitoring. Relevance values can be changed on the basis of the frequency of use, and utility values on the basis of the duration of use, for example. The monitoring of user actions and automatic collection of metadata can be arranged by means of neural networks, for example.

[0025] Wh n synchronization is n eded, an initial data set is d t rmin d 402. The initial data set is preferably a pr -

determined application-specific set. Th user may also add data units to or remove them from the initial data set. Next, metadata associated with the initial data units of the initial data set are retrieved 403, i.e. the links from th initial data units ar defined.

[0026] According to an embodiment of the invention, metadata can be modified 404 according to application or situation. An application- or situation-specific transform function can b used for weighting differ nt data units differ ntly to provide synchronization profiles such as those referred to above. The transform function refers particularly to application- or situation-specific coefficients for the relevance and utility values of the different data units. The transform function is applied to the links between the data units, and the transformed relevance and utility values are then used at later stages (405). This embodiment provides an advantage in that the data units can be weighted differently for different purposes and situations, but employing as small memory space as possible.

[0027] It is also possible to exclude 404 some of the data units indicated in the metadata already before the expected gained utility values are calculated. The exclusion may be based on a minimum value set for utility and/or relevance, in which case only related data units of the initial data set that exceed the minimum value qualify as candidates for the selection data set. When a minimum relevance value is applied, a high-relevance link or a short path can be preferred over long paths of low relevance. If relevance is assigned a high minimum value, the impact of high utility value can be reduced in the selection of data units. For example, a minimum value set for utility allows the synchronization of data units easily obtainable by other means (and thus providing low utility), such as telephone numbers, to be prevented. Another possible exclusion criterion is path length, which allows data units that are too far away from the initial data set to be excluded. In addition, the exclusion method in step 404 allows limit values to be set, whereby expected gained utility values of all data units included in the metadata do not need to be calculated and compared. This speeds up the selection process and reduces the processing capacity needed in the equipment implementing the method. The minimum values applied in the exclusion can also be application-specific, in which case they vary according to the purpose of use.

[0028] The metadata (and the modification and/or exclusion, if any, of step 404) provide related data units associated, one way or another, with the data units of the initial data set. The relevance and utility values denoted by the paths leading to the related data units are preferably used for calculating 405 expected gained utility values E(g) for them. The expected gained utility values obtained for the different data units are compared 406. The data unit with the highest expected gained utility value is added 407 to the selection data set. When a new data unit is added to the selection data set, the routine checks 408 whether an end criterion determined into the data system in advance is met. The end criterion may be, for example, exceeding the maximum size set for the data to be synchronized; exceeding the maximum number of data units; or the non-attainment of minimum expected gained utility value (i.e. there are no data units left which would exceed the minimum value of expected relevance). If the end criterion is not met, the routine proceeds by adding 407 a new data unit to the selection data set.

[0029] Once the end criterion is met, the selection data set, which typically comprises related data units defined according to the initial data set and the end criterion, is synchronized 409. In this connection, the changes made to the selection data set since the last synchronization event can be checked and the changed data units, or at least data about the changes, can be sent to the other party involved in the synchronization. It is to be noted that as a modification to the above description, only the relevance values can be used for selecting the data units (in the comparison 406 or as an end criterion 408). The data selection according to steps 401-408 described above can be carried out in one device participating in the synchronization, in some of the devices, or in all of them.

[0030] The selection data set can be synchronized using any synchronization method. The synchronization may be carried out using a protocol based on the SyncML standard, although the scope of application of the invention is not restricted thereto. According to the SyncML standard, a synchronization session is first initialized in step 409 to select the database to be synchronized. A SyncML client device (TE) comprises a Sync Client Agent executing the SyncML protocol. The client agent may send the SyncML server (S) a SyncML message (Client Modifications) containing information about the changes made to the selection data set since the last message was sent. The SyncML server comprises a Sync Server Agent, which controls the synchronization, and a Synchronization Engine, and it usually waits for the client's initiative for the synchronization. The SyncML server synchronizes the data, i.e. analyses the changes made to the selection data set and harmonizes the data units (makes the necessary additions, replacements and deletions). The SyncML server then sends the client device a Server Modifications message which comprises the information about the changes made to the selection data set since the last synchronization message from the server S. Although simple, the above example serves to illustrate synchronization based on the SyncML standard.

[0031] It is also possible to use a modified SyncML protocol, in which case the data to be synchronized can be selected during the initialization of the synchronization session. According to a preferred embodiment of the invention, it is also possible to define during the synchronization session whether the TE and the S support the adaptive synchronization of the preferred embodiment. In that case the TE uses the initialization message to request the adaptive synchronization type for use, the synchronization type b ing provided with a specific Alert code in the SyncML standard. If the S supports adaptive synchronization, the routine may proceed according to steps 402-408 described above to

select the selection data sets in the synchronization client devic TE and th synchronization server S. When the TE has determined th selection data set, it sends th modifications (Client Modifications) that have taken place since the last synchronization session to the synchronization server S. Th TE may also send additional requir m nts relating to th d termining of the selection data set, for example that a particular data unit must be included in th set, which the serv r S must take into account when selecting th s lection data set. TE prefer nces and oth r data relating to adaptive synchronization may be transmitt d in a Meta elem nt and in an EMI field, for exampl . Th S s lects (402-408) its selection data set in a similar manner. The server S preferably carries out the selection such that at least the data unit modifications sent by the TE are taken into account. Alternatively, it is possible that the S informs the terminal TE about the selection data set it has selected prior to the synchronization. This, however, causes increased delay and adds to the amount of data to be transferred.

[0032] The S harmonizes the data units in the selection data set it has selected on the basis of the modifications sent by the TE and those made into the database (DB) synchronized by the server S. After the harmonization, the S sends the modifications (Server Modifications) that have taken place in the selection data set since the last synchronization session to the TE. On the basis of the modifications, the TE modifies the data units in its memory MEM. According to an embodiment, the TE may send information about the initial data set and other preferences, if any, during the initialization to the server S, which selects the selection data set on the basis of the metadata and the initial data set.

[0033] Figure 5 shows a method according to a second preferred embodiment of the invention in which the metadata can be used also for excluding data units of the initial data set. Metadata, which can be updated in the above described manner and which comprises relevance and utility information, is collected 501 into the system. The relevance and utility values of the data units can be changed, even if the data units concerned were in the initial data set. When a need arises to carry out synchronization, an initial data set is determined 502. Next, at least metadata associated with the initial data units of the initial data set are retrieved 503, i.e. the links between the initial data units are defined.

[0034] The importance of the initial data units with regard to other initial data units is calculated 504. This can be achieved experimentally by removing one data unit at a time from the first data set and by determining, on the basis of the metadata, the expected gained utility value to be obtained if a data unit is added. The expected gained utility values calculated for each initial data unit are compared 505. The initial data unit with the highest expected gained utility value is added 506 to the selection data set. When a new initial data unit is added to the selection data set, the routine checks 507 whether the end criterion determined into the data system in advance is met. The end criterion may be for example the maximum size set for the data to be synchronized, the number of the initial data units, or the non-attainment of the minimum value set for the expected gained utility value. If the end criterion is not met, the routine proceeds by adding 506 the new initial data unit to the selection data set. When the end criterion is met, the initial data units in the selection data set can be synchronized 508. This allows the least relevant initial data units to be removed from the initial data set.

[0035] The embodiment of Figure 5 provides an advantage in that it allows initial data units that have typically been determined in the initial data set on a relatively permanent basis to be placed into an order of relevance and only the most relevant initial data units to be synchronized. The functions shown in Figures 4 and 5 can also be combined, in which case the remaining initial data units are considered to provide the initial data set (step 402) and thus instead of entering step 508, the routine may proceed through step 403 to assess the relevance of the data units related to the initial data units.

[0036] The user's action can be monitored and the metadata updated 501 on the basis of the use of a data unit. For example, the terminal TE may be arranged to monitor the use of audio data files stored in the terminal. When an audio data file has been played, it can be marked for removal, added to the initial data set and replaced by a new audio data file in the next synchronization session. This can also be achieved by changing the relevance and/or utility value to indicate that it is relevant to synchronize an audio data file marked for removal by the audio application. Consequently, an embodiment is provided which allows to determine data units to be removed and to replace the data units during the next synchronization by a new data unit of a similar type.

[0037] It is also possible to apply the method such that data units remaining outside the selection data set after the end criterion has been met are automatically removed. For example, the relevance and utility values of an audio track of a specific music type can be modified according to user behaviour such that instead of being replaced by new ones, the audio tracks of the music type are removed. Similarly, outdated contact information or electronic mail messages can be deleted with this method.

[0038] Data amount can be used as an end criterion in steps 408 and 507. In that case the size of the selection data set is always checked after a new data unit has been added. When a predetermined size limit is reached, the synchronization of the selection data set may begin. According to a preferred embodiment, it is also possible to synchronize data units (or information relating to modifications made to them) one at a time, starting from the data unit that is the closest to the initial data unit. Wh n the predetermined maximum size limit for the data to synchronized is reached, the synchronization is interrupt d. The terminal TE may also s nd th synchronization server S a messag when th

maximum size limit is exceeded so that the S no longer s nds data units for synchronization. In this embodiment, the selection data set is s lected during the synchronization, unlike in Figures 4 and 5. The embodiment's advantages appear in cases wher th size of the data units is not known, the calculation f the size of the data units requires a large processing capacity, or th server d es not know the memory space availabl at th terminal.

[0039] According to a furth r embodiment, data unit size is also tak n int account in the comparison (steps 406 and 505). The ratio of the expected gained utility value E(g) (or the gained utility valu g) to th data amount can be calculated for the data units. The data unit having the highest E(g) per kilobyte is selected (407, 506) into the selection data set. This allows smaller data units to be preferred over larger ones. However, the comparison must be defined such that a small data unit of low relevance is not preferred over a large data unit of high relevance. This can be accomplished for example by applying a logarithm of data unit size, instead of size, in the comparison.

[0040] According to yet another embodiment, the user interface UI;SUI can also be used for inquiring the user about the need for synchronizing one or more data units (before step 409 or 508). This embodiment is useful when large data units are concerned and mainly when the synchronization is to be carried out with a terminal, which has a very limited storage capacity.

[0041] The above-described embodiments are typically applied at the synchronization server S, which selects the selection data set to be synchronized, and, thereby, has an effect on the amount of data to be sent to the terminal TE, which typically has relatively limited memory resources. The present method can also be used in the terminal TE for selecting a selection data set, the modifications made to the set being informed to the synchronization server S. Usually the number of data units added to the terminal TE by the user is fairly small, and thus all new data units (or other modifications made at the terminal TE) can be easily synchronized. However, if savings in time or in transfer costs are to be aimed at, the above solution can also be used to limit the amount of data to be transmitted from the terminal TE for synchronization.

[0042] In server-to-terminal synchronization, different values (relevance, utility) are preferably used in the metadata or in other criteria related to the selection of the data units than in terminal-to-server synchronization. At the server S side, the purpose may be to limit the required memory space (for the TE), whereas the aim at the terminal TE may be to save the processing resources needed for the comparison and selection of the data units. An embodiment of the solution of the invention provides various profiles (with different metadata or different exclusion/end criteria) for different transfer situations. Fast synchronization can be determined for expensive transfer links (through public mobile communications networks) to only synchronize particularly important data units. Full synchronization can be carried out in a local area network of a company, for example.

[0043] Figure 6 further illustrates the initial data set and the selection data set. The initial data set 60 defined with a dotted line comprises four data units with links that illustrate their relationships with other data units. The circles in Figure 6 illustrate all data units which according to the metadata links are in some way associated with the initial data set 60. A dashed line 61 defines the selection data set to be synchronized, obtained by employing the method of the invention. As already described above, one data unit at a time is preferably added to the selection data set 61, the data units that are closest to the initial data units being typically the most important ones as well. It should be noted that the selection data set 61 does not comprise all the data units of the initial data set, i.e. the method illustrated in Figure 5 has been used. Figure 6 further shows a so-called pre-excluded set, defined with a continuous line 62. Expected gained utility values have been calculated for the data units in the set 62, which is selected using the exclusion of step 404. A data unit with a too low relevance value, for example, has been left outside the set 62.

[0044] According to an embodiment, a reference user data unit, which is always included in the initial data set and which has links to other data units, is added to the initial data set 60. The user data unit itself is not a subject of synchronization, but it defines the data units that are to be taken into account when the selection data set is selected.

[0045] It is apparent to a person skilled in the art that as technology advances, the basic idea of the invention can be implemented in various ways. The invention and its embodiments are therefore not restricted to the above-described examples but they may vary within the scope of the claims.

Claims

1. A method for selecting a data set to be synchronized from databases of a data system, **characterized in that** the method comprises the steps of

maintaining in the data system metadata representing the relationships between data units for the purpose of selecting the data set to be synchronized, the metadata comprising at least information on the relevance of the data units with regard to one another;

retrieving metadata associated with at least one initial data unit of a first data set in response to a need to synchronize the first data set;

s lecting a second data set for synchronization, the data set comprising at least on data unit which, on the

basis of said metadata, is most relevant to the initial data unit.

2. A method according to claim 1, **characterized by**
selecting one data unit at a time into the second data set in the order of relevance;
checking the size of the second data set after a new data unit has been added; and
initiating synchronization with the second data set in response to a predetermined size limit having been reached.

3. A method according to claim 1 or 2, **characterized by**
selecting into the second data set only the data units which exceed one or more predetermined exclusion criteria, such as the minimum relevance value.

4. A method according to any one of the previous claims, **characterized in that**
said metadata further comprises utility information representing the utility provided by at least one data unit associated with an initial data unit in said metadata, either directly or through other data units, provided that the initial data unit has been selected.

5. A method according to claim 4, **characterized by**
determining numerical values representing relevance probabilities and utilities between the initial data units and other data units for the metadata;
forming links between the data units in said metadata, the links being associated with the numerical values of at least relevance and utility;
multiplying the relevance values of the links along at least one path originating from the initial data set and leading to other data units;
selecting the utility value of the last link leading to another, separate data unit to be used as the utility of that data unit;
calculating an expected gained utility value for each one of the other data units by multiplying the utility value by the relevance value;
comparing the expected gained utility values of the different data units; and
selecting at least one data unit with the highest expected gained utility value into the second data set.

6. A method according to any one of the preceding claims, **characterized by**
updating said metadata on the basis of user actions.

7. A method according to any one of the preceding claims **characterized by**
adding application-specific metadata to the data system in response to the adoption of a new application; and
retrieving the metadata associated with the at least one initial data unit as required by the application.

8. A method according to any one of the preceding claims, **characterized by**
determining into the data system situation-specific metadata for at least two different synchronization situations; and
selecting the metadata associated with the at least one data unit as required by the synchronization situation.

9. A method according to any one of the preceding claims, **characterized by**
determining the expected gained utility value for initial data units in the first data set by experimentally adding initial data units, one by one, to the first data set, and
selecting into the second data set one or more initial data units the adding of which provides the highest expected gained utility value.

10. A method according to any one of the preceding claims, **characterized by**
the data system comprising at least one synchronization client device and synchronization server;
sending a request for selecting a data set in accordance with the method from the synchronization client device to the synchronization server during the initialization of the synchronization session;
selecting second data sets in the synchronization client device and the synchronization server in accordance with the method;
sending the modifications that have taken place in the second data set since the last synchronization session from the synchronization client device to the synchronization server; and
sending the modifications that have taken place in the second data set since the last synchronization session

from the synchronization server t the synchronization client devic .

11. A synchronization syst m comprising means for synchronizing the data of at l ast two databases, **haracterized in that** the synchronization system further comprises:

means for maintaining metadata representing th relationships between data units, the metadata comprising at least information on the relevance of the data units with regard to one another;
means for retrieving the metadata associated with at least one initial data unit of a first data set in response to a need to synchronize the first data set;
means for selecting a second data set for synchronization, the second data set comprising at least one data unit which, on the basis of the metadata, is most relevant to the initial data unit.

12. A synchronization system according to claim 11, **characterized in that** the synchronization system further comprises:

means for selecting one data unit at a time into the second data set in the order of relevance;
means for checking the size of the second data set after a new data unit has been added; and
means for initiating synchronization with the second data set in response to a predetermined size having been reached.

13. A synchronization system according to claim 11 or 12, **characterized in that**
said metadata also comprises utility information representing the utility provided by at least one data unit associated with an initial data unit in said metadata, either directly or through other data units, provided that the initial data unit has been selected.

14. A synchronization system according to claim 11, 12 or 13, **characterized in that** the synchronization system further comprises:

means for taking application-specific metadata in use in response to the adoption of a new application; and
means for retrieving the metadata associated with at least one initial data unit as required by the application.

15. A synchronization device comprising means for sending modifications made to a data set to be synchronized of at least one database to at least one second party involved in the synchronization, **characterized in that** the synchronization device further comprises:

means for storing metadata representing the relationships between the data units, the metadata comprising at least information on the relevance of the data units with regard to one another;
means for retrieving metadata associated with at least one initial data unit of a first data set in response to a need to synchronize the first data set;
means for selecting a second data set for synchronization, the second data set comprising at least one data unit which, on the basis of the metadata, is most relevant to the initial data unit.

16. A synchronization device of claim 15, **characterized in that**
the metadata also comprises utility information representing the utility provided by the at least one data unit associated with an initial data unit in said metadata, either directly or through other data units, provided that the initial data unit has been selected.

17. A computer software product for controlling a synchronization device, **characterized in that** the computer software product comprises program code which, when executed in the synchronization device, causes the synchronization device to
store metadata representing the relationships between data units for the selection of a data set to be synchronized, the metadata comprising at least information on the relevance of the data units with regard to one another;
retrieve metadata associated with at least one initial data unit of a first data set in response to a need to synchronize the first data set;
select a second data set for synchronization, the second data set comprising at least one data unit which, on the basis of th metadata, is most relevant to the initial data unit.

18. A computer software product according to claim 17, **haracterized in that**
the metadata also comprises utility information representing the utility provided by the at l ast one data unit associated with an initial data unit In said metadata, either directly or through oth r data units, provided that th initial data unit has be n selected.

BNSDOCID: <EP 1267283A2 I >

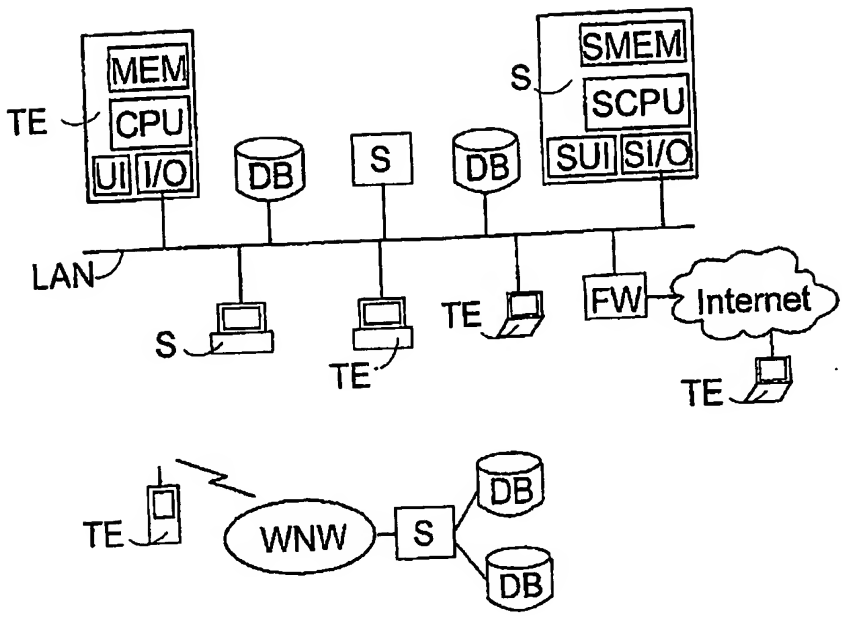

Fig. 1

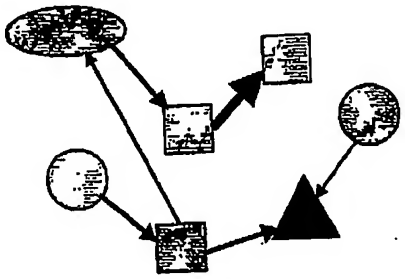

Fig. 2

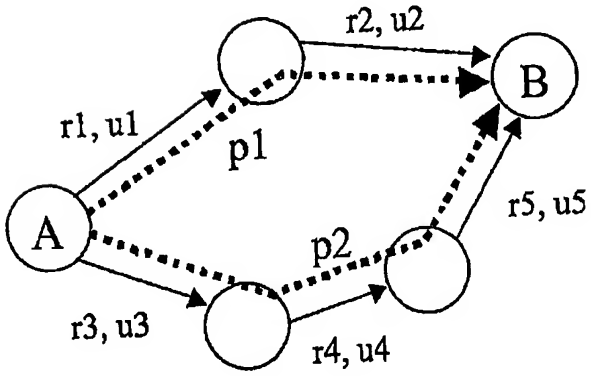

Fig. 3

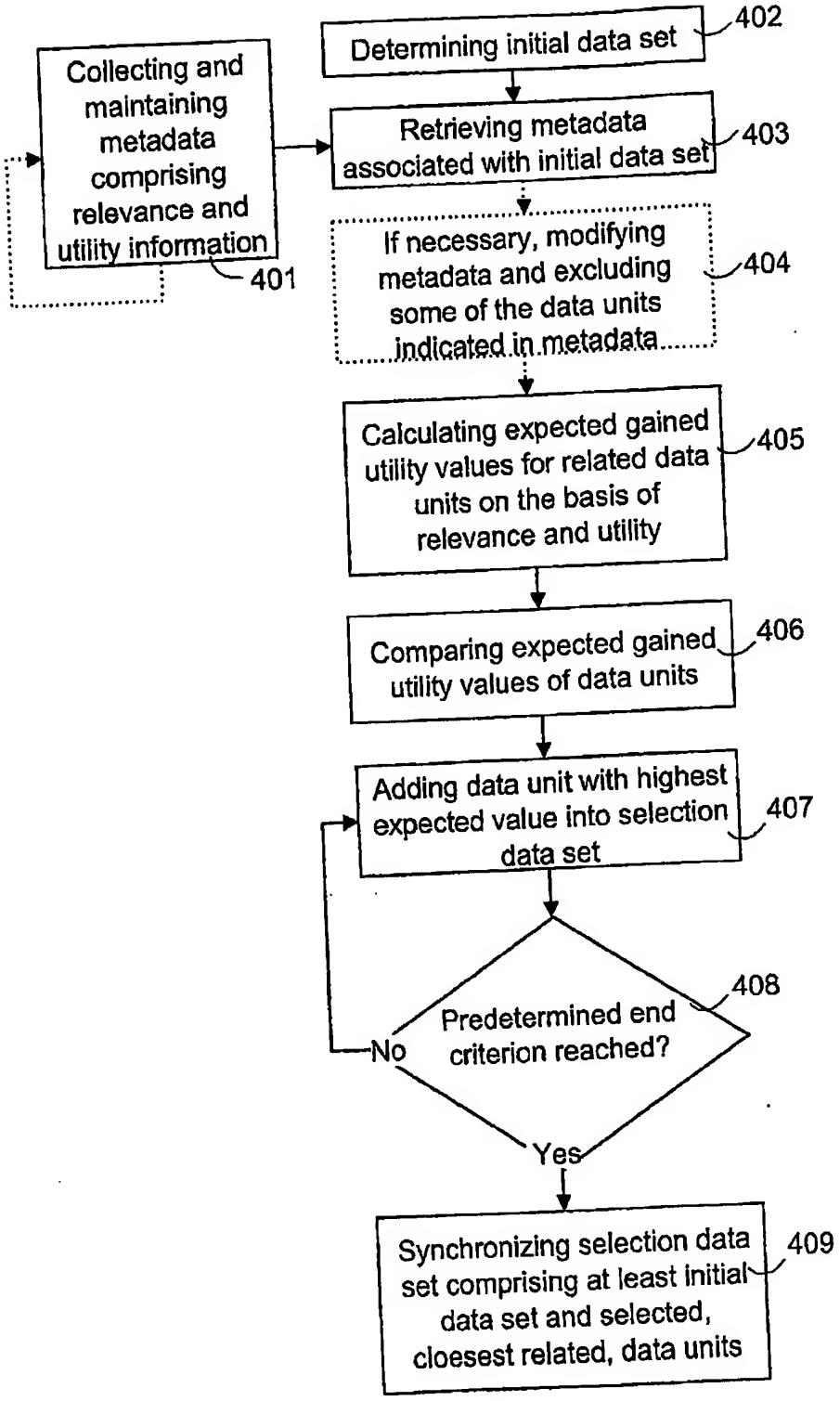

Fig. 4

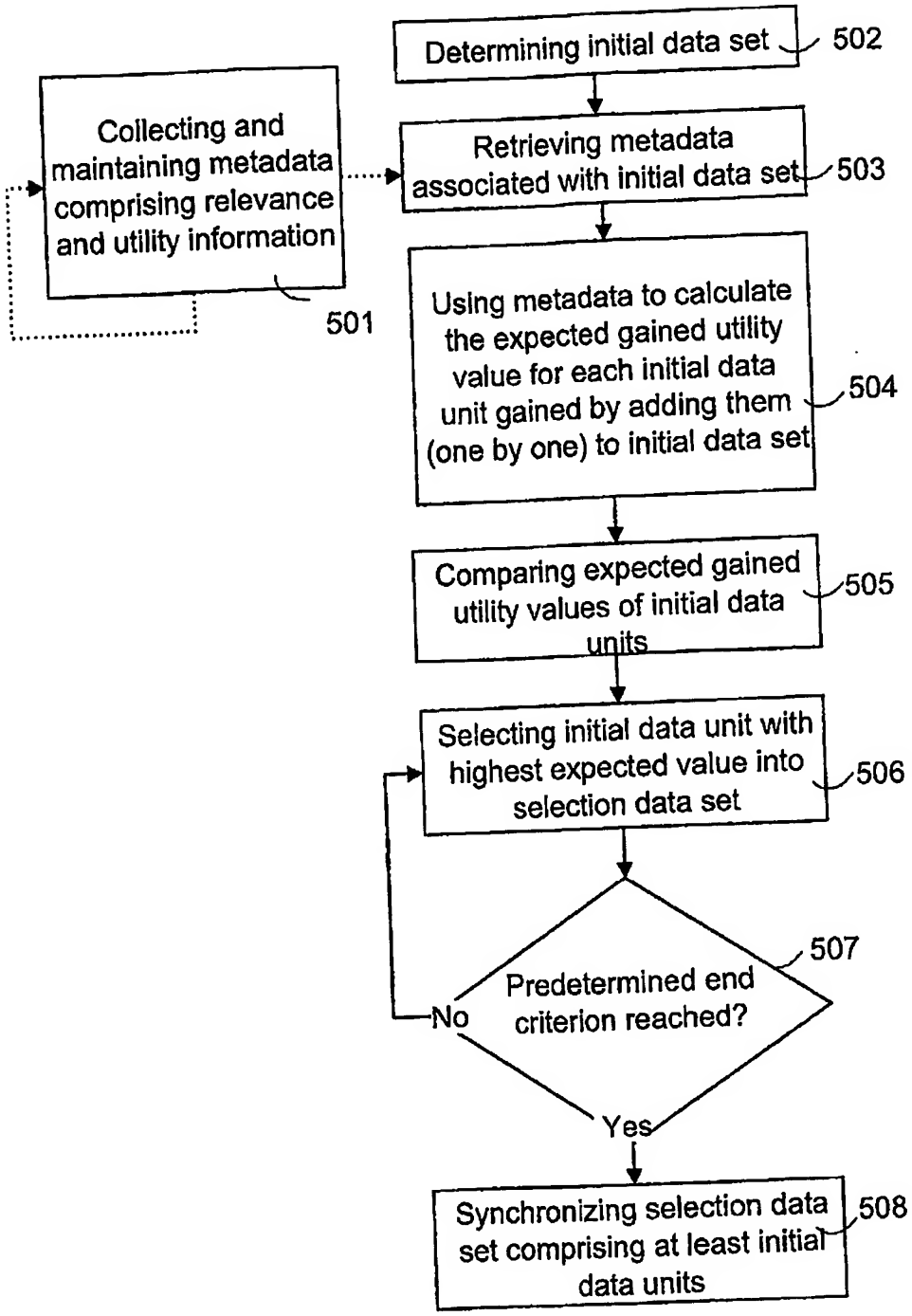

Fig. 5

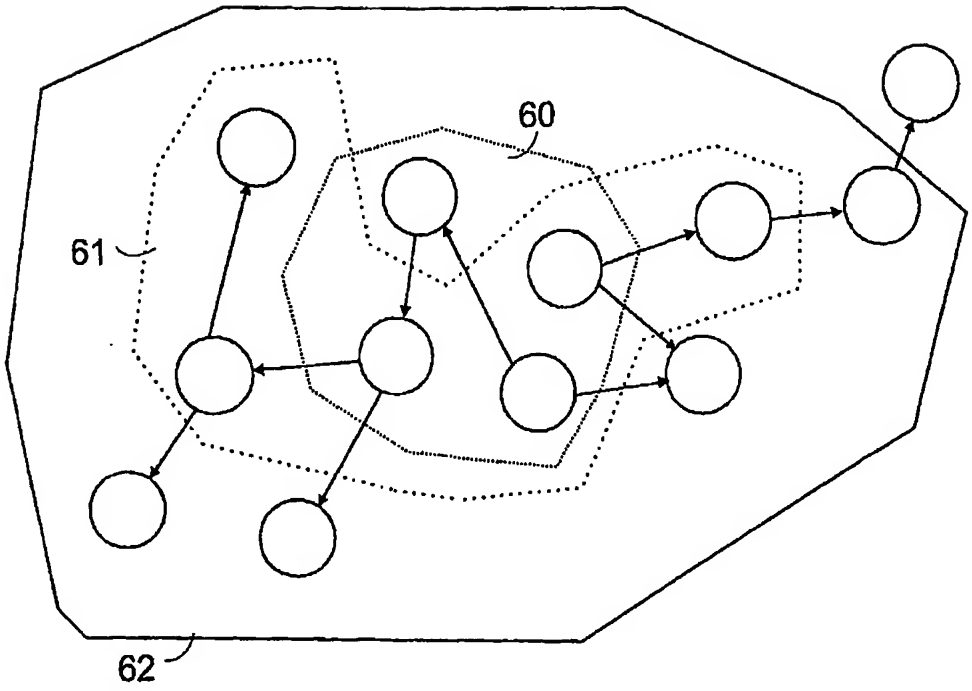

Fig. 6